Weekly Report

2018.1008-2018.1014

1.This Week

Deep Learning Power Grid Project:
1.We finished training the 2000 node dataset and the result is not appreciated. The overfitting problem comes again. We can reach a 97% accuracy on the training set but the accuracy of the validation set can only reach a highest of 57%. We increase the number of samples and reduces the learning rate and did everything we can. But the model cannot learn more. The problem may be the features itself (changes are minor and local in such a large scale grid). We made the following attempts for this problem:

(1) We try to extract more powerful time series features to perform the training as the following paper 1 and 4. So we can reduce the size of each sample and makes the training more efficient.
(2) We try to preprocess the training samples and distinguish faults happened in distribution system and transmission system to narrow down the number of class labels and transform the classifier into parallel classifiers. We are now using information entropy to distinguish between the two but the result is not good.

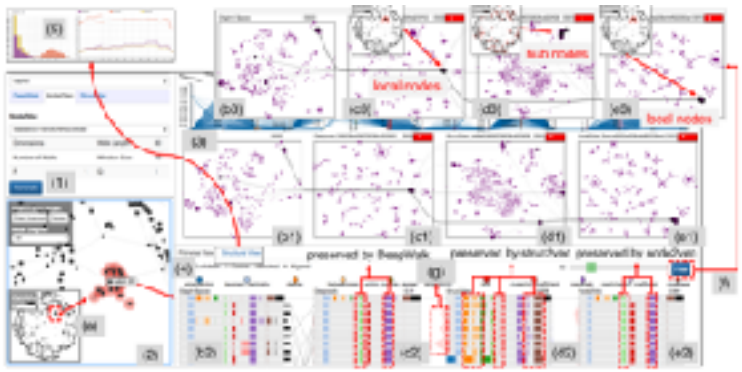
We prepare to make the following attempts:
(1) Use the second vector of tensor CP decomposition as the feature vector to classify the faults. Because the first vector contains the most general information of all nodes and the second is the global space minus the first one so that it remains the amplified features.

Working Hour: (except nap and eat time)
8-9 hours / week day
8 hours on Sunday
Total Working Hour this week: 50.5 hours.

Other
1.Revise RSATree paper for Honghui.
2.Proofreading the monograph.
3.Prepare the materials for the midterm examination.

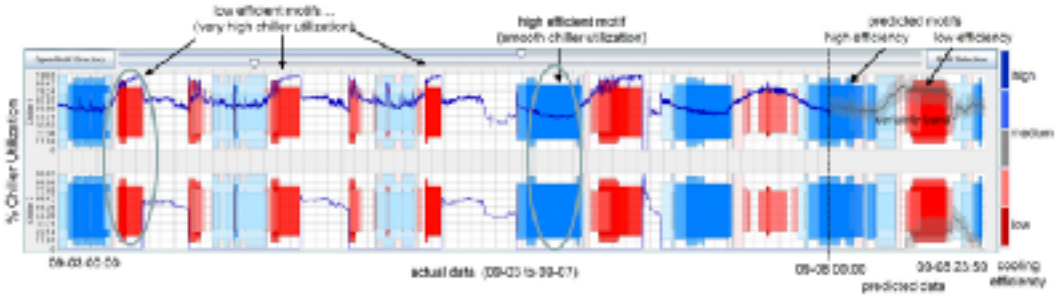
Paper Reading
1.EmbeddingVis: A Visual Analytics Approach to Comparative Network Embedding Inspection (VAST 2018)
This paper compares different graph embedding methods at the cluster, instance, and structural levels. It applies visualization techniques to learn what information

about the original network is preserved in the embedding vectors. Eleven node metrics are computed for each graph embedding space to discover relationships between node metrics and selected embedding vectors



2.Visual exploration of frequent patterns in multivariate time series （Information Visualization）

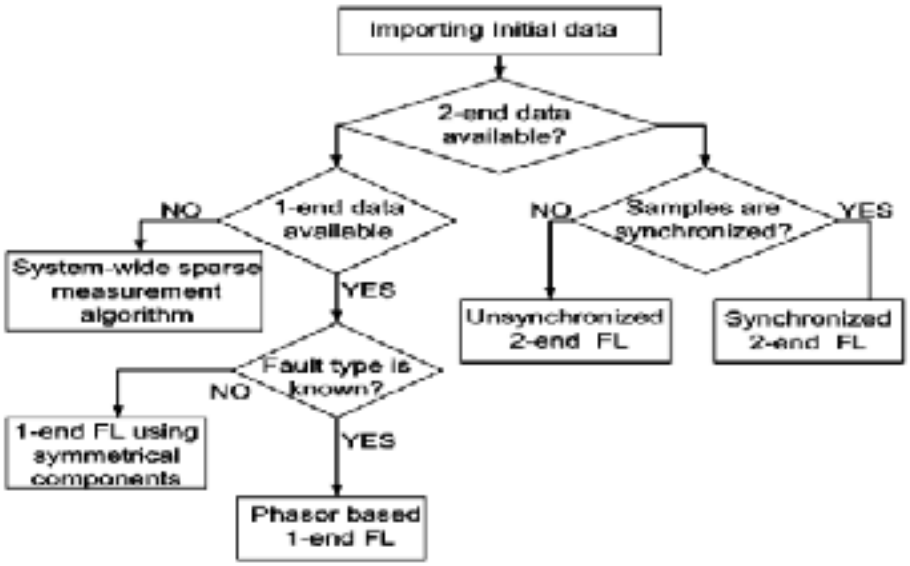
This paper propose a multivariate time series patterns detection algorithm and visual analytics methods to support interactive exploration of the discovered patterns. The pattern detection algorithm first strips off the temporal information, clusters the data, and puts the temporal information back to redescribe the data. An event is defined as a transition in the cluster label. Then events are detected through a Apriori like algorithm. Distortion and merging are supported in the visualization interface to show multi-level motifs (detection patterns) detected in time series.



3.Smart Fault Location for Smart Grids (IEEE smart grid)

This paper integrates heterogenous power grid IED data (real time data) and proposes an optimization flow chart to choose proper algorithm to locate faults to improve accuracy of fault location methods. The flow chart is similar to a predefined decision tree. Fault location methods are separated into distribution system (low base voltage) and transmission system (low base voltage) methods in this paper.

4.The Potential of Household Specific Feature Selection for Analysing Smart Home Time-Series Data (Smart SysTech)



This paper proposes an automatic feature selection algorithm for smart home time series data. The main idea is to generate a feature candidate set and iterativly score the candidates and output a feature set with highest scores finally. The scoring algorithm could be specific with the application domain to provide domain specific features.

5.Dynamic Event Detection Using a Distributed Feature Selection based Machine Learning Approach in a Self Healing Microgrid (IEEE power system)

This paper proposes a feature selection-based distributed machine learning approach to detect dynamic power system events. It uses three features: prominence of local maxima, the width of the local maxima and available frequencies in the time series. Based on these features, a multi-class random forest classifier is used to detect events in time series.

2.Progress

Work	Deadline	Progress
Power grid paper with Deeping learning	12.15	1.Finish training the model on the 2,000 node dataset. 2.Prepare to revise the method.
SQC Paper	-	1.Delayed
WaveLine revision	-	1.Revise the pictures.